



CGAN-rIRN: a data-augmented deep learning approach to accurate classification of mental tasks for a fNIRS-based brain-computer interface

YAO ZHANG,¹  **DONGYUAN LIU,¹**  **TIEN LI,¹** **PENGRUI ZHANG,¹**
ZHIYONG LI,¹ **AND FENG GAO^{1,2,*}**

¹College of Precision Instrument and Optoelectronics Engineering, Tianjin University, Tianjin 300070, China

²Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin 300070, China

*gaofeng@tju.edu.cn

Abstract: Functional near-infrared spectroscopy (fNIRS) is increasingly used to investigate different mental tasks for brain-computer interface (BCI) control due to its excellent environmental and motion robustness. Feature extraction and classification strategy for fNIRS signal are essential to enhance the classification accuracy of voluntarily controlled BCI systems. The limitation of traditional machine learning classifiers (MLCs) lies in manual feature engineering, which is considered as one of the drawbacks that reduce accuracy. Since the fNIRS signal is a typical multivariate time series with multi-dimensionality and complexity, it makes the deep learning classifier (DLC) ideal for classifying neural activation patterns. However, the inherent bottleneck of DLCs is the requirement of substantial-scale, high-quality labeled training data and expensive computational resources to train deep networks. The existing DLCs for classifying mental tasks do not fully consider the temporal and spatial properties of fNIRS signals. Therefore, a specifically-designed DLC is desired to classify multi-tasks with high accuracy in fNIRS-BCI. To this end, we herein propose a novel data-augmented DLC to accurately classify mental tasks, which employs a convolution-based conditional generative adversarial network (CGAN) for data augmentation and a revised Inception-ResNet (rIRN) based DLC. The CGAN is utilized to generate class-specific synthetic fNIRS signals to augment the training dataset. The network architecture of rIRN is elaborately designed in accordance with the characteristics of the fNIRS signal, with serial multiple spatial and temporal feature extraction modules (FEMs), where each FEM performs deep and multi-scale feature extraction and fusion. The results of the paradigm experiments show that the proposed CGAN-rIRN approach improves the single-trial accuracy for mental arithmetic and mental singing tasks in both the data augmentation and classifier, as compared to the traditional MLCs and the commonly used DLCs. The proposed fully data-driven hybrid deep learning approach paves a promising way to improve the classification performance of volitional control fNIRS-BCI.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Brain-computer interface (BCI) technology can bypass the peripheral nervous system and achieve direct interaction between the human brain and external devices (computers and robotic arms, etc.) by decoding the functional activity of the cerebral cortex [1]. Among the neuroimaging modalities applied to BCI, fNIRS has attracted increasing attention due to its advantages as a new non-invasive, portable, low-motion artifact, low-cost, and active measurement-based optical neuroimaging technique with reasonable temporal and spatial resolution, quieter measurement environment and superior motion robustness than functional magnetic resonance imaging (fMRI), stronger environmental robustness and spatial resolution than electroencephalography (EEG)

[2]. Due to the fact that hair induces attenuation of the fNIRS signal, the hairless prefrontal cortex (PFC) becomes an ideal region of interest (ROI) for fNIRS measurements. In general, the user intentionally elicits distinct, repeatable activation patterns in the PFC to control the BCI output by performing different mental tasks, such as motor imagery [3], mental arithmetic (MA) [4], mental singing (MS) [5], and puzzle solving [6], etc. This means that each mental task corresponds to a volitional control state, which is decoded by the classifier to control the BCI output. The motor imagery has been the most commonly used paradigm for BCI control in previous fNIRS-BCI studies [7]. To investigate the suitability of different mental tasks besides motor imagery for BCI control and to improve the discrimination accuracy of activation patterns, two mental tasks for intentional control are performed, namely MA and MS [4]. This is because the MA and MS tasks are the most widely used and robust mental tasks in fNIRS-BCI studies [4,8]. The MA task in brain cognitive activity is an important thinking activity and skill in our daily life, and it is closely related to language and working memory. The MS task, which can also activate PFC, utilizes the positive emotional component of music. When users use and practice the fNIRS-BCI system for a long time to operate peripheral devices, the learning effects and cognitive fatigue across BCI trials can lead to a reduction in accuracy of the mental task. Hence, to overcome these obstacles and to accurately detect and classify the activation patterns of the PFC induced by different mental tasks is crucial for the development of fNIRS-BCI [9].

Feature extraction techniques and classification strategies for fNIRS signals are two important aspects of improving the accuracy of mental tasks, which dramatically affects the performance of BCI applications [10]. In the current research on fNIRS signal classification, many BCI researchers have proposed a large number of classification algorithms, among which machine learning classifiers (MLCs) based on manual feature extraction, such as linear discriminant analysis (LDA) [11], support vector machine (SVM) [12], and adaptive Gaussian mixture model [8,13], etc., have been extensively utilized in fNIRS-BCI due to their low computational cost and relatively better classification performance. Nevertheless, the limitation of traditional MLCs lies in manual feature engineering, involving feature extraction and selection, selection of optimal feature subset, and data dimensionality reduction, which are considered to reduce the accuracy. These inherent bottlenecks also cause many researchers to spend a substantial amount of time on data pre-processing and feature extraction.

The fNIRS signal is a typical multivariate time series with multi-dimensionality and complexity, which makes DLC an ideal alternative to overcome the above dilemma [10,14]. The fully data-driven DLC is an automatic feature extractor and classifier that automatically extracts deep features and discriminates subjects' intentions from brain activation patterns captured in PFC regions. The hemodynamic parameters reflecting the metabolism of cortical tissues are changes in the concentrations of oxygenated hemoglobin ($\Delta[\text{HbO}]$), deoxygenated hemoglobin, and total hemoglobin [7]. Hence, we can feed one of the above three hemodynamic signals (the $\Delta[\text{HbO}]$ signal is commonly used as its more pronounced amplitude changes can more sensitively reflect changes in regional cerebral blood flow) or multiple ones (stacked by temporal or spatial dimensions) to the DLC for fNIRS signal classification.

The purpose of using DLC in the volition-controlled fNIRS-BCI is to improve discrimination accuracy for multi-class mental tasks. In the current research on DLCs applied to fNIRS signal classification, this paper categorizes DLCs into univariate time series (UTS) based DLC, multivariate time series (MTS) based DLC, and spectrogram (i.e., time-frequency representation) based DLC according to the representation of the fNIRS signal fed to the DLC. In the UTS-DLCs studies, Hiwa *et al.* performed a one-dimensional convolutional neural network (CNN) with a convolutional layer and a pooling layer to classify the UTS-fNIRS signal for each sampling location. The sampling with the highest accuracy was defined as the critical ROI, and the generated label was used to discriminate the subject's gender [15]. In another work, Yoo *et al.* used a long short-term memory (LSTM) network to classify three mental tasks, involving

MA, mental counting, and puzzle solving, in eight healthy subjects to reduce processing and classification time, with a maximum accuracy of 83.3% compared to SVM and LDA [6]. In a recent work, Asgher *et al.* increased the number of commands and adopted two DLCs, namely CNN and LSTM, to classify four-level mental workload states, and the results demonstrated that DLCs enhanced the accuracy compared to traditional MLCs (SVM, k -nearest neighbor, and artificial neural network) with an average accuracy of 87.45% and 89.31%, respectively [16]. UTS-DLCs paving the way for the development of portable or wearable fNIRS-BCI systems with a few sampling locations due to their high-speed data processing capabilities for the rapid positioning and discrimination of critical ROI of neural activation in the brain. However, due to the complexity and stochasticity of neural activation patterns, it also reveals that the flaw of UTS-DLCs is that their input data do not have spatial distribution information and only use a single sampling location with the highest accuracy to discriminate the differences in spatial patterns, which leads to a significant degradation of accuracy [7].

How to make full use of the temporal and spatial distribution information of fNIRS signal to more accurately discriminate complex hemodynamic pattern changes is an urgent problem for DLCs applied to task classification in fNIRS-BCI. Hence, the crucial solution to the above problem lies in the development of MTS-DLCs. In the MTS-DLCs studies, Trakoolwilaiwan *et al.* proposed a CNN model for classifying three tasks of left- and right-handed motor execution and resting state of eight subjects, and the results showed that the accuracy of the proposed CNN was improved by 6.49% and 3.33% compared to the SVM and artificial neural network methods, respectively [14]. The model is a variation on the CNN architecture commonly used for image recognition. The input of the network is a two-dimensional (2D) matrix composed of multi-sampling locations time series signals, but the size of its convolution kernels is not the conventional size, its width covers the whole sampling location. The convolution kernel moves only in the direction of the time dimension. The drawback of this CNN is that it only extracts the shallow features of the entire spatial distribution information and cannot obtain the multi-scale spatial features, and cannot unmix the contribution of samplings. In a recent work, Ma *et al.* proposed the LSTM-Inception DLC for accurate classification of left- and right-handed motor imagery [7]. The inception network architecture is proposed by Google for image recognition, which improves the recognition performance by broadening the structure of the network [17]. The LSTM-Inception method uses a bottleneck layer to directly reduce the spatial dimensionality. LSTM and Inception extract critical features for long and short intervals of one-dimensional time series, respectively. This method is also ineffective in obtaining multi-scale spatial feature information.

In the studies of spectrogram-based DLCs, inspired by EEG-DLCs, Janani *et al.* used short-time Fourier transform to convert UTS into 2D spectrogram data for CNN classification [10]. Although this spectrogram representation of fNIRS signal improves the classification performance, the increased data dimensionality renders the computational complexity higher. To sum up, most of the DLCs reviewed above focus on time series properties of the fNIRS signal without fully considering the spatial distribution information. Besides the above three DLCs based on signal representation, Saadati *et al.* proposed a spatial CNN to classify fNIRS signals [18]. The spatially represented signal has only spatial distribution information but not temporal correlation. Ghonchi *et al.* used a deep recurrent-CNN (RCNN) to classify synchronous EEG-fNIRS signals [19]. The three convolutional layers and two LSTMs in the RCNN extract the temporal and spatial features of the synchronous signals, respectively. The structure of RCNN fully considers the spatio-temporal characteristics of synchronous signals, but does not make special structural adjustments and designs according to the spatio-temporal diversity, and does not have adaptive multi-scale feature extraction. Therefore, an elaborate DLC is desired for multi-task and accurate classification of mental tasks in fNIRS-BCI.

The other non-negligible problem is that DLC has an inherent bottleneck in that vast amount of training data and computational resources are required to enable deep networks to be adequately trained to prevent overfitting [20]. However, the collection of substantial-scale, high-quality labeled fNIRS data is complex and expensive, or even impractical, when comprehensive considering factors such as time, labor, and material costs, stability of the measurement system, and the comfort of the subjects during long-term continuous measurements. Data augmentation is a critical technique for tackling the issue of small sample datasets and the long-tail effect of data (i.e., unbalanced data distribution) for DLC [21]. However, traditional data augmentation approaches in image recognition, involving scaling, rotation, cropping, shifting, and other geometric transformations are not applicable to multi-channel fNIRS data that do not have a single-target [22]. Similarly, in the EEG-BCI study, empirical modal decomposition approaches were used to augment the data for improving accuracy and reducing the time spent on rehabilitation training [23]. With the development of artificial intelligence, generative adversarial networks (GANs) are used to generate artificial samples to augment the training dataset and improve the accuracy [24,25]. Based on the above, this study addresses two target concerns, namely insufficient training data and accuracy improvement. We herein propose a novel data-augmented DLC for classifying MTS-fNIRS signals, which combines conditional generative adversarial network (CGAN) based data augmentation and rIRN-based DLC. The convolution-based CGAN is utilized to generate class-specific synthetic fNIRS signal to augment training dataset, allowing DLC to be fully trained. The rIRN network architecture is characterized by serial multiple spatial-FEMs and temporal-FEMs, where each FEM performs deep and multi-scale feature extraction and fusion, and the ResNet extracts global features and prevents gradient disappearance.

To demonstrate the feasibility and superiority of the proposed CGAN-rIRN approach for improving the accuracy of mental tasks in fNIRS-BCI, the classification performance of the rIRN approach was compared with that of traditional MLCs (SVM and LDA) and commonly used DLCs (Inception-ResNet (IRN), CNN, and back propagation neural network (BPNN)) for MA and MS mental tasks in eight subjects. In addition, the effect of different levels of augmented training datasets on classification accuracy was also further analyzed.

2. Methods

The overall architecture of the CGAN-rIRN model is shown in Fig. 1. The CGAN-rIRN is a hybrid with two deep learning models. The convolution-based CGAN model generates synthetic fNIRS signals for augmenting the training dataset of the DLC, while the rIRN model is an elaborate DLC for accurate classification of mental tasks in fNIRS-BCI.

For the input data of the hybrid network model, the raw light intensity signals are converted into filtered $\Delta[\text{HbO}]$ signals using the modified Beer-Lambert Law (MBLL) [11] and band-pass filtering (BPF) [26]. The filtered $\Delta[\text{HbO}]$ signals are randomly shuffled and divided into training dataset, validation dataset, and test dataset in the split ratio of 4: 1: 1. The training dataset is fed to CGAN for generating class-specific synthetic fNIRS signal. The real and synthetic fNIRS signals are converted into graphics (i.e., stored as grayscale images) and combined into an augmented training dataset, which is then fed into the rIRN-DLC model as an input layer.

2.1. Convolution-based CGAN model

Goodfellow *et al.* first presented the concept of GAN [27], which is broadly utilized in the field of computer vision, such as face generation [28], due to its powerful generative power. In recent years, GAN has also been adopted for data augmentation in the fNIRS-BCI field [24]. In that study, the problem of gradient vanishing was solved by using Wasserstein GAN (WGAN). However, the WGAN requires a separate run for each category of fNIRS signal to augment the training dataset for that category of signal. If the number of fNIRS signal categories is too high, the WGAN is too cumbersome to use and increases the computational complexity.

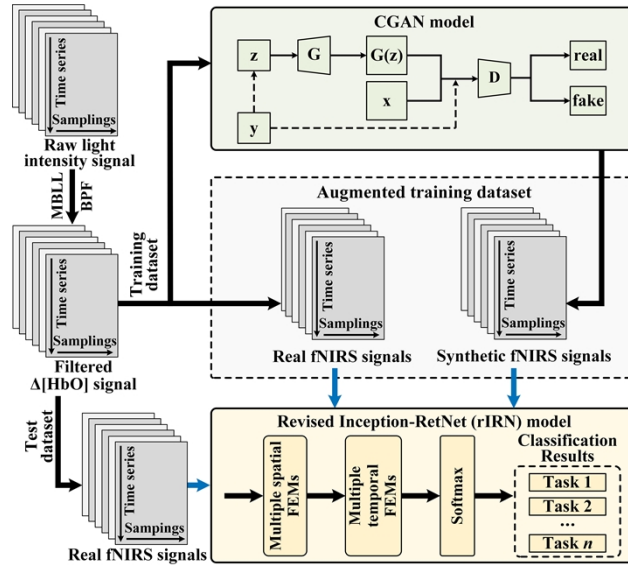


Fig. 1. The overall architecture of the CGAN-rIRN model.

To interpret the artificial samples generated by the GAN and to avoid the computational complexity of multiple runs of the algorithm, we herein propose a convolution-based CGAN model to generate category-specific synthetic signal. The CGAN is a probabilistic generative model with conditional constraints, which has an additional conditional variable \mathbf{y} as auxiliary information (category labels) for both the generator and the discriminator compared to GAN. Our proposed CGAN architecture consists of two separate networks that are trained simultaneously, a generator that generates synthetic fNIRS signal with the identical distribution and diversity as the category-specific signal, and a discriminator that attempts to discriminate the synthetic fNIRS signal from the real fNIRS signal, as illustrated in Fig. 2. The value function (objective function) of CGAN can be expressed as the following maximum and minimum problem in Eq. (1):

$$\min_G \max_D [V(G, D)] = \min_G \max_D \{ \mathbb{E}_{\mathbf{x} \sim \mathcal{P}_{real}(\mathbf{x}), \mathbf{y} \sim \mathcal{P}(\mathbf{y})} [\log D(\mathbf{x}|\mathbf{y})] + \mathbb{E}_{\mathbf{z} \sim \mathcal{N}(0,1), \mathbf{y} \sim \mathcal{P}(\mathbf{y})} [\log(1 - D(G(\mathbf{z}|\mathbf{y})|\mathbf{y}))] \}, \quad (1)$$

where $\mathbf{x} \in \mathbb{R}^{T \times C}$ is the real signal (i.e., single-trial $\Delta[\text{HbO}]$ signal) with probability distribution $\mathcal{P}_{real}(\mathbf{x})$, T is the number of sampling points in the single-trial fNIRS data, and C is the number of sampling locations. In the present study, T is 160 (i.e., $40 \text{ s} \times 4 \text{ Hz}$) and C is 10. The \mathbf{y} is the category label with probability distribution $\mathcal{P}(\mathbf{y})$, which is attached to the generator and discriminator through the embedding layer. $\mathbf{z} \in \mathbb{R}^{1 \times 100}$ is the random noise obeying the standard normal distribution $\mathbf{z} \sim \mathcal{N}(0, 1)$, which is fed into the generator as an input layer. $G(\cdot)$ is the generator which accepts noise \mathbf{z} and embedded category labels \mathbf{y} to generate synthetic fNIRS signal, $G(\mathbf{z}|\mathbf{y})$ is the synthetic signal generated under the guidance of condition \mathbf{y} . $D(\cdot)$ is the discriminator that represents the probability of judging the data as real under condition \mathbf{y} . \mathbb{E} denotes mathematical expectation. The settings of the loss functions for $G(\cdot)$ and $D(\cdot)$ can be found in Gauthier's study [28].

The generator is a transposed convolutional network consisting of a fully connected layer (FCL) with 7680 neural nodes and three 2D transposed convolutional layers (T-Conv2D) with filter numbers of 32, 16, and 1, respectively. The convolution kernel for the whole generator is set to 3×3 , with a stride (the sliding step of the filter) of 2 except for the last T-Conv2D layer which has a stride of 1. The output of the generator is the synthetic signal with a size of $T \times C$.

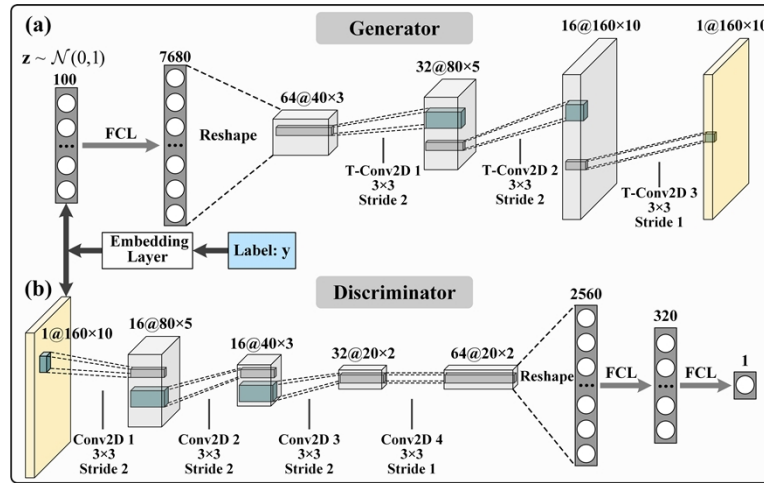


Fig. 2. The architecture of the convolution-based CGAN model: (a) the generator, (b) the discriminator.

and fed into the discriminator. The discriminator is a convolutional network consisting four 2D convolutional layers (Conv2D) with filter numbers of 16, 16, 32, and 64, respectively, and two FCLs with 320 and 1 neural nodes, respectively. Except the convolution kernel stride of the last Conv2D layer is 1, all other strides are 2. The activation functions of all T-Conv2D and Conv2D layers are set to the leaky rectified linear unit (LReLU) with a slope of 0.2, and the feature maps obtained after T-Conv2D and Conv2D layers are batch normalized before activation.

In the training of CGAN, the input size of the generator is a \mathbf{z} -vector with 100-dimension, the input size of the discriminator is $T \times C$ matrix of real fNIRS signal (\mathbf{x}) and synthetic fNIRS signal ($G(\mathbf{z}|\mathbf{y})$), and the training batch size is 4. The update of the weight coefficients of the generator and the discriminator alternates in each epoch, updating one of them while the other remains unchanged. The number of epochs is set to 500. The weight coefficients of both networks are randomly initialized, and the gradient descent optimization algorithm is an adaptive moment estimation optimizer [29] with a learning rate of 0.0002, where parameters β_1 and β_2 are set to 0.5 and 0.999, respectively.

2.2. Revised Inception-ResNet model

The proposed rIRN model is inspired by the Inception-ResNet network and combined with the spatial and temporal characteristics of the fNIRS signal, as shown in Fig. 3. The overall architecture of the rIRN model is shown in Fig. 3(a).

The spatial-FEMs consists of two sub-modules, namely spatial-FEM-A and spatial-FEM-B modules, where each sub-modules comprises a spatial Inception-ResNet module and a spatial reduction module. The spatial-FEM-A sub-module is shown in Fig. 3(b). The spatial Inception-ResNet-A module consists of one-dimensional convolution filters with lengths of 3, 5, and 7 for extracting spatial features at different scales, and it only slides along the spatial-dimension to compute the convolution and the output dimension remains unchanged. Both the max-pooling layer and the last layer of each feature extraction branch in the spatial-reduction-A module have a stride of 2 to achieve spatial dimensionality reduction. The spatial-FEM-B module has a similar structure to the spatial-FEM-A module, with the difference that its filters have lengths of 2 and 3, and the number of all its convolution filters is 32, while that of spatial-FEM-A module is 16.

The temporal-FEMs consists of three sub-modules, which are constituted by temporal-FEM-A, -B and -C modules. The temporal-FEM-A sub-module is illustrated in Fig. 3(c). The structure of

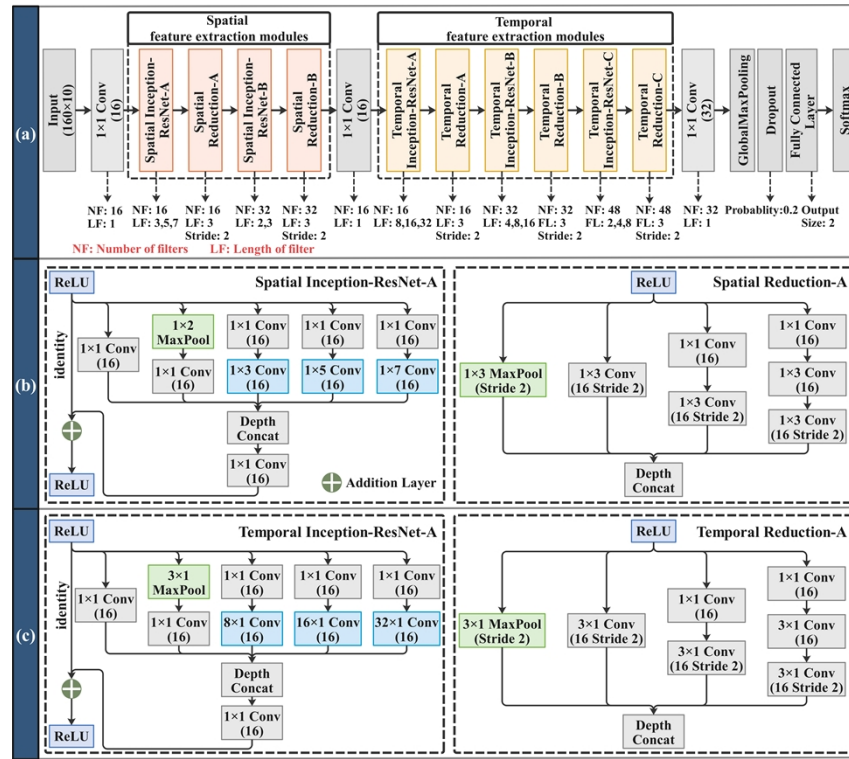


Fig. 3. The architecture of the rIRN model. (a) Flow chart of the overall architecture of the rIRN. (b) Spatial-FEM-A and spatial dimensionality reduction A modules. (c) Temporal-FEM-A and temporal dimensionality reduction A modules. ReLU denotes rectified linear unit. “Depth Concat” indicates depth concatenation layer.

each sub-module of the temporal-FEMs is similar to that of the spatial-FEMs. They differ in the length and sliding direction of the filter, where the filter for temporal-FEMs only moves along the direction of the temporal-dimension. Since the dimensionality of the time series in single-trial fNIRS signal at high sampling rate is much larger than that of the spatial sampling locations, the length of the convolution filter in multi-scale extraction of time-dimensional features is longer than that of the spatial one. Therefore, the filter lengths of temporal Inception-ResNet-A sub-module are 8, 16, and 32, respectively, those of -B sub-module are 4, 8, and 16, and those of -C sub-module are 2, 4, and 8. As the temporal-dimension of the data is reduced after processing by each sub-module of temporal-FEMs, the filter length of the later sub-module is shorter than that of the previous one. The number of all filters for the temporal Inception-ResNet-A, -B, and -C sub-modules are 16, 32, and 48, respectively. The structure of the temporal reduction module is similar to that of the spatial one.

In the training process of the rIRN, the validation dataset is employed to examine the state and convergence of the rIRN. Moreover, a random grid search method was used to explore the different hyper-parameters of the rIRN to determine the optimal values with the best classification performance [14]. The final hyper-parameters of the rIRN are set to a batch size of 32, a maximum epoch of 100, an L2 regularization of 0.0001, and a learning rate of 0.001 or 0.0001 which is appropriately reduced to the lower one as the size of the training dataset increased [30]. The adaptive moment estimation optimizer is used for the gradient descent optimization algorithm. The CGAN-rIRN model is built using MATLAB R2020b (MathWorks, USA) on a

64-bit operating system in Microsoft Windows 10. The training and testing conditions for the CGAN and rIRN models are the same. The CGAN and rIRN models are trained and tested on a server with 128GB RAM, two Intel Xeon Gold 6138 CPU @ 2.00 GHz, and a NVIDIA(R) GeForce RTX 2080Ti GPU.

3. Experiments

3.1. Participants and data acquisition

Eight healthy right-handed participants (mean age: 25.4 ± 2.0 years, 2 males and 6 females) were recruited from the students at Tianjin University for the experiments. None of the subjects reported a history of any psychiatric, neurological, or brain disorder. All participants were asked to refrain from drinking alcohol, smoking, and caffeinated beverages for 6 h prior to each data collection procedure, and to relax previous to the experiment to stabilize blood flow. The study was conducted with informed consent and received ethical approval from Tianjin University.

The fNIRS data were acquired using a lab-made portable continuous-wave diffuse optical tomography system. The system adopts a phase-locked photon counting technique to achieve fully parallel measurements at dual-wavelengths (785 nm and 830 nm) [31]. According to the international 10-20 system, four laser diode source-pairs and four photomultiplier tube detectors formed a single-lattice arrangement with 20 measured channels of raw light intensity (i.e., 10 sampling locations \times 2 wavelengths per source-pair), which were then placed in the PFC using a homemade flexible headband and covered with optode positioning points FP1 and FP2 to collect fNIRS signals, as exhibited in Fig. 4(a). The source-detector distance was set to 30 mm and the system sampling rate was set to 4 Hz, which ensures a high signal-to-noise ratio of the measured data.

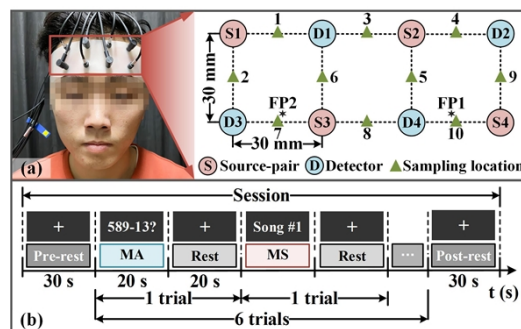


Fig. 4. Acquisition of fNIRS data. (a) Experimental source and detector configuration. (b) Experimental paradigm.

3.2. Experimental protocol

The participants were asked not to move any part of the body, and performed MA and MS mental tasks according to the instructions displayed on the screen during the data recording. In the MA mental task, participants were asked to perform a sequence of simple mathematical calculations in which a small number (between 9 and 15) was repeatedly subtracted from a randomly generated three-digit number. In the MS mental task, participants silently rehearsed self-selected Chinese song fragments that they felt would elicit within them a strong and positive emotional response, but they were asked to choose the songs that they liked. They were instructed to try to feel the emotions of the songs rather than simply memorize the lyrics or tune. In each trial, the self-selected Chinese song was different for each participant, and its randomly generated from the subject's self-constructed song library.

The schematic diagram of the experimental paradigm is depicted in Fig. 4(b). Each subject was required to perform 24 data collection sessions on the same day, with a diagram of one of the sessions shown in Fig. 4(b). A data collection session consisted of a 30 s pre-rest period (i.e., baseline period), six 40 s trials, where the first 20 s period of each trial was an MA or MS mental task, followed by a 20 s rest period (alternating between MA and MS trials), followed by a 30 s post-rest period. Data collection for all subjects was completed within one week. The fNIRS optodes were not removed from the subjects during all data collection session. The total number of trials collected per subject for each class of mental task was 72. Hence, the total number of trials collected per subject was 144.

3.3. Signal pre-processing and feature extraction

The differential pathlength factor (DPF) in MBLL was set to $DPF_{785\text{ nm}} = 6.35$ and $DPF_{830\text{ nm}} = 5.88$ [32]. A 5th-order zero-phase Butterworth digital BPF with cut-off frequencies of 0.018 and 0.3 Hz was used to eliminate baseline drift, low-frequency oscillations, global systemic noise, respiration, and high-frequency physiological interference (heartbeat) of the raw $\Delta[\text{HbO}]$ signal. Ma *et al.* stated that the respiratory rate for adults is about 0.4 Hz [7]. Since the subjects recruited were all adults (graduate students in college), the low-pass cut-off frequency of BPF was chosen to be 0.3 Hz. Then, a 3rd-order least squares smoothing filter is adopted to the BPF-filtered data to mitigate the oscillations of Mayer waves and the interference of random noise on the fNIRS signal. The baseline calibration of the filtered $\Delta[\text{HbO}]$ signal for each session of each subject uses the data from the baseline period under that session. The coefficient of variation was used to assess the quality of the measured data [8]. If there were large motion artifacts in the measured data, the data from this session were discarded and required to be remeasured [33]. The quality assessment and re-measurement was done immediately after the session.

The filtered $\Delta[\text{HbO}]$ data were split into samples according to the single-trial period, in which each sample consisted of 40 s of $\Delta[\text{HbO}]$ data, with 20 s task period and 20 s rest period. We extracted the mean and slope features of $\Delta[\text{HbO}]$ data and normalized them to [0, 1] as inputs to MLCs and BPNN, respectively. In our previous study, it was revealed that these two features can yield better classification performance than the commonly used statistical features [8]. Due to the small number of sampling locations under the optode distribution, there is no need to use feature selection methods for dimensionality reduction. This is also to compare the classification performance of DLCs and MLCs more fairly.

3.4. Controlled classification methodologies

The rIRN model was trained utilizing an augmented training dataset consisting of a real fNIRS dataset and a synthetic fNIRS dataset. The data for the augmented training dataset is a fixed-size graphics of 160×10 without feature selection and normalized to [0, 1]. When the dimensionality of the input data varies due to the number of sampling points and spatial sampling locations, the rIRN model can adjust the length of the corresponding convolution filters and the number of FEMs to achieve the best classification performance. The accuracy was used as an evaluation metric to quantitatively evaluate the ability of the classifier to accurately discriminate between mental tasks [25], as defined below:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \quad (2)$$

where TP is the number of true positives, FP is the number of false positives, TN is the number of true negatives, and FN is the number of false negatives. To demonstrate the efficiency and superiority of the proposed CGAN-rIRN in enhancing the accuracy of mental tasks for fNIRS-BCI, we compared traditional MLCs including LDA, SVM with linear kernel function and regularization parameter of 1, and commonly used DLCs comprising IRN, CNNs, and BPNN.

3.4.1. Inception-ResNet model

We compare the proposed rIRN model with the existing IRN model. The architecture of the IRN takes reference from the Inception branch in the LSTM-Inception network proposed by Ma *et al.* [7]. It has six Inception modules, where every three Inception modules have one residual connection. Each Inception module has one-dimensional convolution filters with lengths of 10, 20, and 40, which are used to extract the feature information of time series with different lengths. The difference between rIRN and IRN is that the former is carefully designed to extract features at different scales in the spatial and temporal dimensions, respectively, while the latter is only designed to extract feature information in the temporal dimension and does not capture deep spatial feature information. In addition, rIRN has both spatial and temporal dimension reduction modules. The size of the filter for each FEM of the rIRN is reduced accordingly as the depth of the network increases in this dimension, while the size of the one-dimensional filter in IRN is fixed and does not change with the deepening of the network.

3.4.2. Convolutional neural networks

The architecture of the control CNN classifier is referenced from the CNN model with two convolutional layers (abbreviated as CNN-2 L) proposed by Trakoolwilaiwan *et al.* [14]. The CNN-2 L consists of two convolutional layers, where the size of the filters in each layer is 3×10 and 3×1 , respectively. All the filters shift only in the temporal-dimension of the input data. The number of filters in each convolutional layer is 32 and 64, respectively. Each convolutional layer is followed by a batch normalization layer, a ReLU layer, and a max-pooling layer with stride of 2. The Dropout layer behind the last max-pooling layer has a probability of 0.2 to drop out the input elements. Three FCL layers have 128, 16, and 2 hidden nodes, respectively. The output layer is a softmax function. The optimal hyper-parameters in the CNN-2 L are set to a learning rate of 0.001, a batch size of 16, and a maximum epoch of 100. To compare with the proposed rIRN more fairly, we also use a CNN model with six convolutional layers (abbreviated as CNN-6 L), which makes the two models similar in model complexity. The number of filters for each convolutional layer of CNN-6 L is 16, 16, 16, 32, 32, and 32.

3.4.3. Back propagation neural network

The BPNN classifier has a hidden layer, and the optimal number of nodes in the hidden layer is set according to the following modified empirical formula [34]:

$$n_{hid} = \lceil \log_2(n_{inp}) \rceil + a, \quad (3)$$

where n_{hid} and n_{inp} are the number of nodes in the hidden layer and input layer, respectively. The a (set to 2) is an integer within the range [1,10]. $\lceil \cdot \rceil$ indicates that the ceiling is a rounding to positive infinity operator. The BPNN uses a variable learning rate to accelerate the convergence of the model, which is set as follows at the current iteration number t :

$$\eta(t) = \eta_{\max} - t(\eta_{\max} - \eta_{\min})/t_{\max}, \quad (4)$$

where η_{\max} and η_{\min} denote the maximum and minimum learning rates of 0.01 and 0.00001 respectively, and t_{\max} indicates the maximum iteration number of 1000. The stochastic gradient descent method was used in the updating the weights and biases of the BPNN. The input to the BPNN was the manually extracted features. To better evaluate the classification performance of each classifier, the average accuracy obtained from three runs of the classifier on the test dataset was used as the final evaluation of the model.

3.5. Statistical analysis

A paired-samples *t*-test with a significance criterion of 0.05 are performed using OriginPro 2022b software (OriginLab Inc., Northampton, Massachusetts, USA) to statistically analysis the classification accuracy of MA and MS mental tasks for all participants.

4. Results

4.1. Qualitative evaluation of the synthetic fNIRS signal

The quality of the synthetic fNIRS signal needs to be evaluated as it has a significant impact on the training effectiveness of the classification model. To evaluate the quality of synthetic fNIRS signal generated by the proposed convolution-based CGAN model (CGAN-CON), we compared it with the full connection-based CGAN model (CGAN-FUL) in both qualitative and quantitative terms in a fair manner. Both the generator and the discriminator of the CGAN-FUL model are fully connected networks. The generator has three hidden layers with 256, 512, and 1024 nodes, respectively. The number of nodes in each hidden layer of the discriminator is 1024, 512, and 256, respectively. The optimal hyper-parameter settings for the CGAN-FUL are batch size of 4, learning rate of 0.0002, and epoch of 2000.

Figure 5 illustrates graphics (i.e., normalized grayscale images are converted into pseudo-color images for visualization) of the real fNIRS signal collected by participant 1 in performing the MA and MS mental tasks and the synthetic fNIRS signal generated by the two CGANs. As observed from Fig. 5, the synthetic fNIRS signal generated by CGAN-CON has a crisp, high-resolution graphic compared to the real fNIRS signal and is indistinguishable to the naked eye. However, the quality of the synthetic fNIRS signal generated by CGAN-FUL is poor, with certain artifacts and distortions.

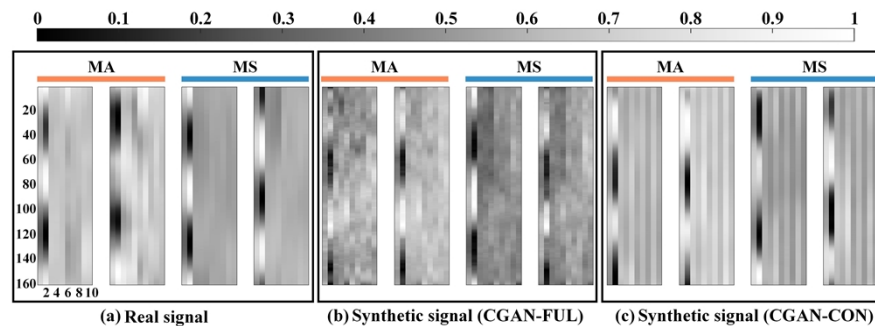


Fig. 5. Graphical display of the real and synthetic fNIRS signals for MA and MS mental tasks of participant 1: (a) real fNIRS signal, (b) synthetic fNIRS signal generated by CGAN-FUL, (c) synthetic fNIRS signal generated by CGAN-CON. The color bar indicates the normalized grayscale value.

To further compare the similarity of the synthetic fNIRS signals generated by the two CGANs with the real fNIRS signal, the two middlemost sampling locations in the measurement region of the source-detector configuration (i.e., sampling locations #5 and #6) were selected to visualize the five trials averaged fNIRS signal to qualitatively analyze the quality of the synthetic signals. Figure 6 shows the real and synthetic fNIRS signals averaged over five trials for representative sampling locations of participant 1. The results demonstrate that the time-varying forms of the synthetic fNIRS signals generated by CGAN-CON for MA and MS were consistent with those of the real fNIRS signals. In contrast, the synthetic signals generated by CGAN-FUL had larger fluctuations and were more different from the real fNIRS signal. In addition, during the stimulus period, significant hemodynamic responses were observed between the synthetic signals

generated by the two CGANs and the real fNIRS signals for samplings #5 and #6 of MA and sampling #5 of MS. Although the amplitude of the hemodynamic response of the real fNIRS signal for sampling #6 of MS is small, CGAN-CON can generate a synthetic signal with high similarity to the real fNIRS signal, while no significant hemodynamic response was observed for the synthetic signal of CGAN-FUL.

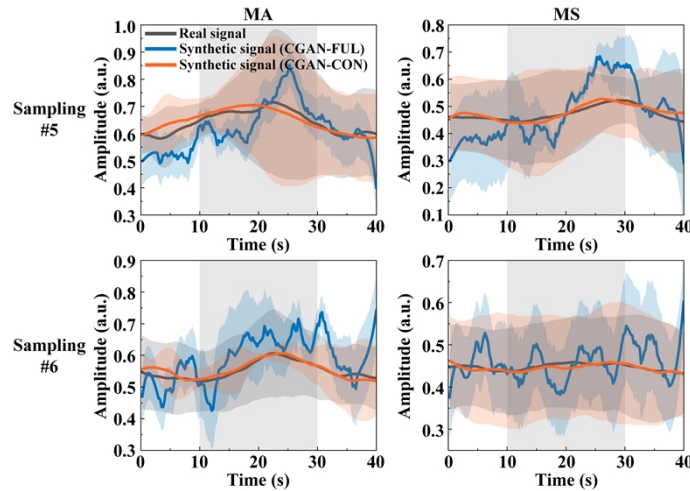


Fig. 6. Examples of five-trial averaged real and synthetic signals for sampling locations #5 and #6 of participant 1 in the MA and MS mental tasks. The shades of the same color around the curves represent the standard deviation. The gray highlighted rectangles indicate the stimulus periods.

4.2. Quantitative evaluation of the synthetic fNIRS signal

We also quantitatively evaluated the quality of the synthetic fNIRS signals generated by the proposed CGAN-CON with that of CGAN-FUL. Three commonly used evaluation metrics, including maximum mean discrepancy (*MMD*) [35], structural similarity index measure (*SSIM*) [36], and peak signal to noise ratio (*PSNR*) [36] were chosen to quantitatively evaluate the signal quality generated by two CGANs. The *MMD* metric is widely used in transfer learning to measure the distance between the distributions of two different random variables [37]. It is utilized to measure the distance between the distribution of real and synthetic fNIRS datasets for a particular category of task. The smaller the *MMD* distance, the smaller the discrepancy between the two distributions. The mean *MMD* was calculated by randomly selecting 40 samples from each participant's real and synthetic dataset under each category of mental task.

The *SSIM* is a multi-scale structural similarity method for evaluating the closeness of the mean luminance, contrast, and correlation of the real and synthetic fNIRS signals. The higher the similarity between the graphics of the two fNIRS signals, with *SSIM* being closer to 1. The *PSNR* is a widely used full-reference graphical quality evaluation metric. It is also utilized to evaluate the quality of generated signal under a category-specific task. Obviously, the larger the *PSNR* value, the smaller the difference between the real and synthetic fNIRS signals, further indicating that the quality of the synthetic signal is better. 10 randomly selected real-synthetic sample pairs were used to fairly calculate the mean *SSIM* and *PSNR* values for each subject under a specific category. The 10 averaged *SSIM* and *PSNR* values were the final evaluation results for this category of fNIRS signal.

The results of the quantitative comparison of the synthetic signals of MA and MS generated by the two CGANs for all participants are shown in Fig. 7(a)-(c) and Fig. 7(d)-(f), respectively.

As can be seen from the results, the mean *MMD* values across all subjects for the synthetic MA (Fig. 7(a)) and MS (Fig. 7(d)) signals generated by CGAN-CON are lower than those of CGAN-FUL. For the synthetic MA signals, the mean *SSIM* (Fig. 7(b)) and *PSNR* (Fig. 7(c)) values of CGAN-CON were higher than the results of CGAN-FUL. For the synthetic MS signal, the results of CGAN-CON (Fig. 7(e)-(f)) are similarly higher than those of CGAN-FUL. There were no significant differences in *SSIM* and *PSNR* metrics between the two CGANs for all subjects in both MA and MS signals, indicating that their generated synthetic data have comparable performance in terms of structural similarity and distortion compared to the real signal, with no superiority or inferiority. Although, compared with the CGAN-FUL, the mean values of the *SSIM* and *PSNR* metrics of CGAN-CON are higher, this does not completely indicate that its performance is better than that of the CGAN-FUL. However, there are significant differences between the two CGANs in terms of *MMD* metrics, further suggesting that the CGAN-CON generated synthetic signal is closer to the distribution of the real signal. Therefore, the superior performance of CGAN-CON over CGAN-FUL in terms of quality of generated signal is supported by the results of the *MMD* metrics. Considering the three metrics together, we are still able to draw the preliminary conclusion that CGAN-CON is superior to CGAN-FUL in generating synthetic signal.

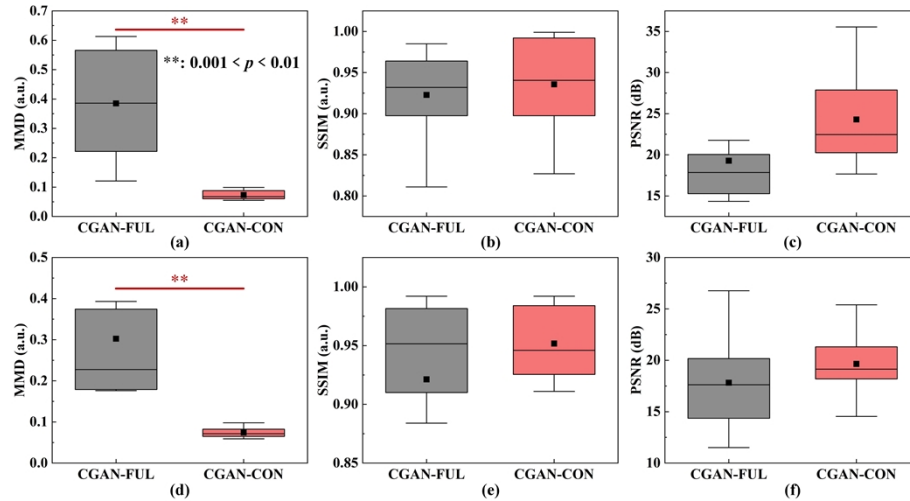


Fig. 7. Quantitative comparison of the synthetic fNIRS signals generated by the two CGAN models. (a)-(c) MA mental task, (d)-(f) MS mental task. Black squares denote mean values.

The combined analysis of above qualitative and quantitative results of the synthetic MA and MS signals both support that the CGAN-CON has better performance than CGAN-FUL to generate high-quality fNIRS signals. Therefore, CGAN-CON was used as a data augmentation method for fNIRS signals. The CGAN in CGAN-rIRN mentioned in this article refers to CGAN-CON by default.

4.3. Effect of different levels of augmented training datasets

To verify the effectiveness and superiority of the proposed CGAN-rIRN method in improving the accuracy of mental tasks, we analyzed the effects of different levels of augmented training datasets on the single-trial accuracy of MLCs and DLCs by varying the amount of augmented training data. The classifiers are trained and tested on each subject separately as a way to perform subject-level decoding. The augmented training dataset consisted of measured original fNIRS signal and augmented fNIRS signal generated by CGAN-CON, and it had the same amount of

augmented data for each category of mental task. The notation “ N ” was used to denote the number of original fNIRS training datasets, where the amount of measured data for both the MA and MS tasks was $N/2$. N was set to 96 for each subject, i.e., $144 \times (4/6) = 96$. We analyzed the effect of different sizes of augmented training datasets with dataset levels of $1.5N$, $2N$, $4N$, $6N$, $8N$, $10N$, $20N$, $30N$, and $40N$ on the accuracy of the classifier, and then compared the classification performance between different classifiers under the same settings. For example, an augmented training dataset with a $1.5N$ level comprises an original dataset of N level and an augmented dataset with of $0.5N$ level.

The single-trial average accuracy across all subjects of each classifier on different levels of the augmented training datasets is presented in Fig. 8. From the results of DLCs in Fig. 8(a), it can be seen that the average accuracies of rIRN, IRN, CNN-2L, and CNN-6L increase with the number of augmented training datasets when the level of augmented training datasets $\leq 2N$, and their maximum accuracies all appear at the dataset level of $2N$. A further observation reveals that when the level of augmented training datasets $> 2N$, their accuracy no longer increases but decreases in fluctuation. However, the data augmentation did not improve the classification performance of the BPNN based on manually extracted features. Comparing the results of CNN-2L and CNN-6L reveals that CNN-6L with a deeper network does not display a better classification performance than CNN-2L. The accuracy of the proposed rIRN was the highest among the above four classifiers at different levels of augmented training datasets. The highest decoding accuracy of rIRN was 92.19% at the dataset level of $2N$.

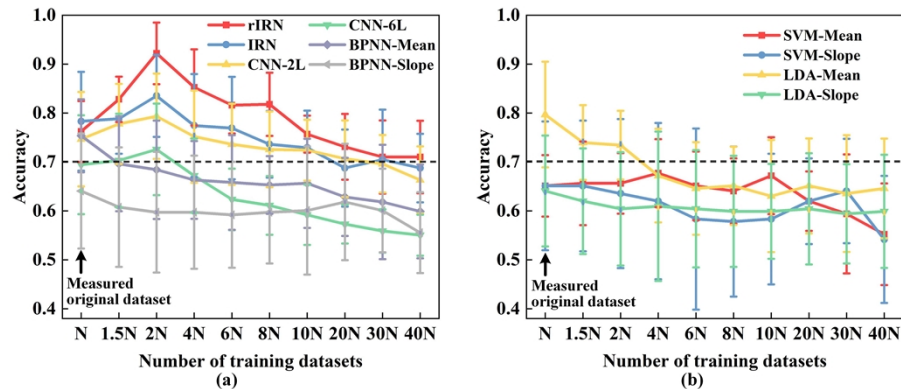


Fig. 8. Effect of the different levels of augmented training datasets on single-trial accuracy for (a) different DLCs and (b) different MLCs. “-Mean” and “-Slope” denote classifiers based on the mean and slope features, respectively. The error bars on each data point indicate the standard deviations. The black dashed line indicates 70% classification accuracy of effective binary BCI communication.

From the results of the MLCs in Fig. 8(b), it can be seen that the proposed data augmentation method only improves the accuracy of SVM-Mean, where the maximum improvement is 3.99% at the dataset level of $4N$. However, the results of the other MLCs show that their accuracies were not effectively improved by the augmented training dataset, especially the LDA-Mean has a significant decrease in accuracy. In addition, we also observe that the accuracy of MLCs was mostly below 70%, except for the LDA-Mean accuracy at the level of augmented training dataset $\leq 2N$. Comparing the results in Figs. 8(a) and 8(b), it can be found that the accuracy of rIRN was higher than that of the control other DLCs and all MLCs at all different levels of augmented training dataset. This also further demonstrates the overwhelming decoding advantage of the proposed rIRN method.

4.4. Statistical analysis of the results before and after data augmentation

The maximum improvement in the accuracy of the proposed data augmentation approach to most DLCs was at the 2N level of the training dataset. Therefore, the results of the average decoding accuracy of each classification algorithm before and after data augmentation at this level and the associated statistical information are illustrated in Fig. 9. As observed from the results, the CGAN-CON-based data augmentation method has indeed improved the average accuracy of the DLCs such as rIRN, IRN, CNN-2L, and CNN-6L by 20.95%, 6.65%, 6.27%, and 4.51%, respectively, compared to the average accuracy based on the original dataset. This result was consistent with the previous findings by Wickramaratne *et al.* [25]. However, the data augmentation did not improve the accuracy of the BPNN. For the results of the MLCs, the data augmentation marginally improves the average accuracy of SVM-Mean by 0.80% for the same level of training dataset, while it did not enhance the accuracy of other MLCs. In the classification results based on the original dataset, LDA-Mean had the highest accuracy of 79.69% among all the classifiers. However, when the original training dataset was augmented by the CGAN-CON, the accuracy of our proposed rIRN was significantly improved to 92.19%. The paired-sample *t*-test yielded significant differences in the mean accuracy across all subjects between the rIRN and the other control classification methods when the training dataset was augmented. Hence, the proposed CGAN-rIRN approach turned out to be an effective and superior classification method in accurately classifying MA and MS mental tasks.

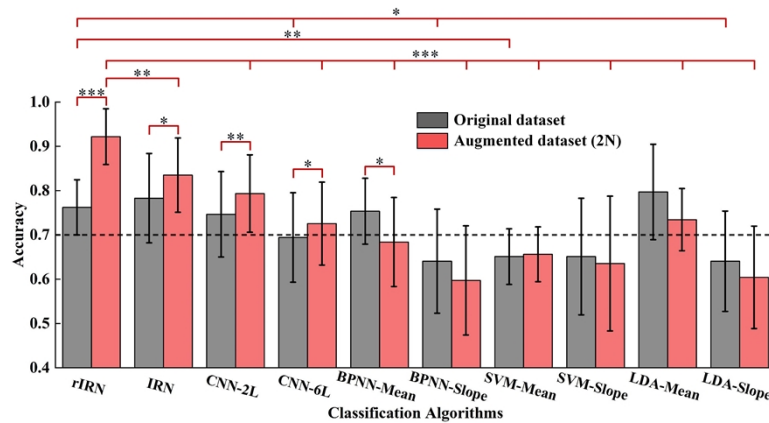


Fig. 9. The average accuracy of each classifier before and after the training dataset augmentation. The augmented dataset is at 2N level. The black and red bars indicate the decoding accuracy based on the original and the augmented training datasets, respectively. Error bars indicate the standard deviation, * represents the significant difference, empty: $p > 0.05$, *: $0.01 < p < 0.05$, **: $0.001 < p < 0.01$, ***: $p < 0.001$.

5. Discussions

5.1. Selection of evaluation method for classification models

Leave-one-out (LOO) [12] and k-fold cross-validation (CV) [14] methods are widely used in the performance evaluation of classification models. The amount of available data and computing resources are the two main concerns when choosing an evaluation method. The LOO-CV retains a larger sample of data but is only suitable for small datasets because it is too costly to compute. The k-fold CV method is usually used for relatively larger datasets. However, it is not suitable for evaluating the deep rIRN on larger augmented datasets, considering computational complexity, resources, and time. In addition, there are limitations to the above two CV methods when using

augmented datasets to evaluate model performance. For the LOO-CV, the distribution of the synthetic data is different from that of the real data, which may lead to a biased evaluation of the synthetic data as a test set and does not fully reveal the true generalization ability of the model on the original dataset. For k-fold CV, a similar problem is faced when dividing the augmented dataset into k-folds for CV. This is because the distribution of data may also be different for each fold, which ultimately affects the accuracy of the assessment results.

In summary, we divide the original signal dataset into a training set, a validation set, and a test set. This is also a common approach in the field of deep learning [38]. The synthetic data generated by the proposed CGAN-CON was only used to augment the training dataset. The biggest advantage of this dataset segmentation method is that it is simple, fast, easy to comprehend and implement for larger datasets with data augmentation. In this approach, the model is constructed and trained using only the augmented training set, the real validation set is used for hyper-parameter tuning and model selection, and finally the real test set is used to evaluate the performance of the final model.

5.2. Overfitting problem

The overfitting problem can lead to a lack of generalization ability of the CGAN-rIRN. The following measures have been taken to mitigate the effects of overfitting problems: (a) Data augmentation. For the limited real training data of rIRN, the CGAN was used to augment the amount of training data to mitigate the overfitting problem. (b) Regularization. We regularize the parameters using the L2-norm to prevent overfitting of the model. (c) Early stopping. We stop training if the validation loss does not decrease 20 times in a row to prevent the model overfitting on the training set. (d) Dropout. We added a Dropout layer to the CGAN-rIRN to randomly discard some neurons in the intermediate layers, thus reducing the possibility of overfitting. We have experimentally validated the effectiveness of these strategies and will continue to explore more effective approaches in future studies.

5.3. Classification results

To validate the effectiveness of the proposed CGAN-CON for data augmentation, we fairly compared it with the CGAN-FUL. The results of the qualitative and quantitative comparison of the synthetic fNIRS signals suggest that the proposed CGAN-CON has better performance than CGAN-FUL in generating stable, diverse, and high-quality fNIRS signals. In addition, previous study by Nagasawa *et al.* showed that the fNIRS signal generated by WGAN has an additional noise component [24], whereas the proposed CGAN-CON generates a high-quality smoothed signal that is highly similar to the real signal, as shown in Fig. 6. Although only a fair comparison between CGAN-CON and CGAN-FUL has been made in this paper, our future study will compare with more data augmentation methods to demonstrate the superiority of the proposed data augmentation method.

The classification results for the MA and MS tasks in Fig. 8 demonstrate that the accuracy of proposed rIRN was higher than that of the control DLCs and MLCs based on each of the different levels of augmented datasets. The results further suggest that the rIRN can significantly improve the accuracy of the mental task. The accuracy of rIRN increases with the number of augmented training datasets when the level of augmented datasets was $\leq 2N$, and it reaches the maximum value of 92.19% at the $2N$ level. This regularity was also present in other DLCs (IRN, CNN-2 L, and CNN-6 L), which suggested that augmenting the real signal with synthetic signal can significantly improve the accuracy of DLCs except for the BPNN based on manual feature extraction. Therefore, data augmentation is an effective strategy to improve the accuracy of DLCs based on automatic feature extraction to classify fNIRS signals, which finding was consistent with the study of Wickramaratne *et al.* [25]. When the level of the dataset was $> 2N$ (i.e., the proportion of augmented data in the augmented dataset was greater than the measured original

data), the accuracy of DLCs decreases with the number of augmented datasets. It is possible that the distributions of synthetic and real signal are closer, but not equivalent. When the number of original signals in the augmented training dataset was less than that of the augmented signals, the DLC learns more about the intrinsic patterns of the synthetic signal and overfits the original signal, which eventually reduces the accuracy of the test data.

In our study, no consistent regularity was found in the effect of CGAN-CON-based data augmentation on the accuracy of MLCs. The accuracy of SVM-Mean was slightly improved, while that of the other MLCs was reduced. In other words, the preliminary conclusion is that data augmentation does not have an enhancing effect on the accuracy of MLCs for the classifying MA and MS tasks. This finding is different from the study by Nagasawa *et al.* where WGAN-based data augmentation provided a large improvement in the classification performance of the SVM [24].

The augmented dataset with $2N$ level has the highest accuracy improvement for DLCs, with rIRN, IRN, CNN-2 L, and CNN-6 L achieving an average accuracy of $92.19 \pm 6.29\%$, $83.53 \pm 8.38\%$, $79.34 \pm 8.73\%$, and $72.57 \pm 9.36\%$, respectively, as shown in Fig. 9. Compared with IRN ($p = 0.86 \times 10^{-2}$), CNN-2 L ($p = 3.35 \times 10^{-4}$), and CNN-6 L ($p = 4.37 \times 10^{-5}$), the average accuracy of rIRN improved by 10.37%, 16.20%, and 27.04%, respectively. Therefore, our proposed rIRN has more significant classification superiority compared to the CNN-2 L proposed by Trakoolwilaiwan *et al.* [14]. The accuracy of CNN-2 L was higher than that of CNN-6 L for both the original and augmented training datasets, which also demonstrates that simply increasing the number of convolutional layers does not necessarily lead to an improvement in accuracy. This finding is consistent with the results obtained by Trakoolwilaiwan *et al.* using CNNs with different structures [14]. Therefore, fine-tuning the structure of deep learning networks was a complex technical problem that heavily relies on the rich experience of the designer. The elaborate design of the network structure and the choice of the optimal hyper-parameters were also important factors affecting the classification performance of DLCs.

The learning rate in the gradient descent algorithm has a significant impact on the accuracy of rIRN. The appropriate learning rate for rIRN was selected based on the size of the augmented dataset. When the level of the augmented dataset is $\geq 20N$, the learning rate was adjusted from 0.001 to 0.0001, which will result in a better accuracy of rIRN. This finding can also be analyzed from the principle of the gradient descent algorithm when using the sum of squared errors as a cost function [30].

5.4. Computational time

In a realistic application scenario, fNIRS-BCI system urgently requires a fast and efficient classification method for real-time high-accuracy classification of fNIRS signals. To validate the real-time classification performance of the developed fully data-driven rIRN-DLC approach, the computation time of the model should be measured. The training times for both the DLC and MLC models were computed with an augmented training dataset at $2N$ level and their hyper-parameters were kept constant as described above. To simulate a real application environment, a single sample was fed into the trained deep learning model with a batch size of 1 to calculate the testing time. The training and testing times were averaged across all subjects.

The GPU computation times for each classification algorithm to classify MA and MS tasks are shown in Table 1. The results show that the training and testing times for DLCs were more time-consuming than that of MLCs. The training and testing times for LDA were slightly higher than those for SVM, respectively. For the training process of the DLCs, the training time of rIRN was 2.09, 19.11, 10.72, and 116.72 times greater than that of IRN, CNN-2 L, CNN-6 L, and BPNN, respectively. For the testing time of all DLCs, the computation time of rIRN was 1.68, 6.08, 4.81, and 473.23 times higher than that of IRN, CNN-2 L, CNN-6 L, and BPNN, respectively. Hence, the training and testing time of rIRN was more time-consuming than that of

the control DLCs. This is due to the deeper and more complex network structure of the rIRN, which is specifically-designed according to the characteristics of the fNIRS signal. Although this designed architecture increases model and computational complexity, but it provides better performance in classification accuracy.

Table 1. The GPU computation time (in s) for each classification algorithm

Time	Classification algorithms						
	rIRN	IRN	CNN-2L	CNN-6L	BPNN	SVM	LDA
Training time	511.25	244.26	26.75	47.69	4.38	5.70×10^{-3}	8.10×10^{-3}
Testing time	9.37×10^{-2}	5.57×10^{-2}	1.54×10^{-2}	1.95×10^{-2}	1.98×10^{-4}	2.68×10^{-5}	4.54×10^{-5}

If a sliding window with a smaller time window is used to feed real-time data to the rIRN, this makes its prediction time shorter and will greatly satisfy the requirements of real-time BCI. Therefore, the model can improve the accuracy, stability, response time, and efficiency of BCI applications. It also provides strong support for real-time control tasks based on fNIRS signals, and promotes the development of BCI technology applications in production and medical fields.

5.5. Future study

The proposed CGAN-rIRN is a pervasive multi-task classification approach that can adaptively adjust the size of the spatio-temporal filters according to the fNIRS settings or recording parameters. In different experimental paradigms and recording parameters, CGAN is able to generate synthetic data with the same dimensional size as the real data from 100-dimensional random noise. The rIRN can adaptively adjust the size of the spatial filter according to the number of channels to achieve multi-scale spatial feature extraction, and can also adaptively adjust the size of the temporal filter according to the sampling rate and the experimental paradigm of the block design to perform multi-scale feature extraction in the temporal dimension. Therefore, CGAN-rIRN can be applied to fNIRS data across diverse conditions.

This study has some limitations that will be tackled in the future study. First, we collected a smaller sample of subjects from the student population in the data acquisition. To ensure relative homogeneity, we should overcome the current difficulties and recruit more subjects from different populations and ages. Second, the cognitive fatigue across trials plays an important role and influence in the development of a multi-task fNIRS-BCI algorithm [39]. In future study, we should add and improve on the following aspects: (a) the subjects' fatigue levels can be systematically quantified by attention tests, fatigue questionnaires, and physiological signal monitoring to assess the subjects' physical status and attention levels, (b) the changes in classification performance at different time periods (e.g., earlier, middle, and later trials) during the experimental process can be analyzed, which can provide a more comprehensive understanding of the performance differences of the rIRN model across trials and effectively circumvent the uncertainty and error caused by fatigue, and (c) compare the changes in classification performance between different tasks and within each task separately to assess the effectiveness of the multi-task algorithm. These expanded experiments can make the conclusions of this study more comprehensive and accurate, and can also provide more valuable references for subsequent studies. Third, the comparison of classification methods is adequate, but also incomplete. In the state-of-the-art classification methods, LSTM [16] and RCNN [19] are also applied in the classification of fNIRS signals. To more fully verify the superiority of the proposed rIRN, these two methods should also be adequately compared with rIRN. Fourth, the CGAN-rIRN has achieved a significant improvement in accuracy for intra-subject decoding mental tasks. However, cross-subject decoding is a future direction for fNIRS-BCI and has been extensively investigated in the fields of fMRI [40] and EEG [41]. In the future study, cross-subject fNIRS decoding will be an important application direction for the rIRN. The leave-one-subject-out cross validation was used to evaluate the classification

performance of the established subject-specific and population-wide classification models [42]. Fifth, transfer learning has also been frequently used in cross-subject decoding [20,43], so combining rIRN and transfer learning for cross-subject fNIRS decoding is also a future study direction for us. Sixth, the proposed CGAN-rIRN is a multi-task classification approach with universality. To demonstrate the versatility and applicability of the rIRN in real-world BCI scenarios, we should really consider classifying a broader range of mental tasks, such as word generation, motor imagery, and tangram puzzle, etc [44]. Seventh, the quality assessment of the synthetic data is comprehensive, but it can also be validated more comprehensively from the perspective of statistical signal characterization. Statistical features, such as power spectrum and autocorrelation function, can be extracted from the synthetic data and the real data for comparison. In addition, we have only used the real and augmented datasets to train the classification model, and also used a separate synthetic dataset to train the classifier to more fully analyze its impact on the classification performance. In the data augmentation, the superiority of the CGAN-CON needs more metrics to make a more comprehensive validation. Eighth, in order to adequately assess the robustness of the rIRN model and the reproducibility of the classification results, more suitable cross-validation methods should be considered.

6. Conclusion

We propose a fully data-driven hybrid deep learning approach, CGAN-rIRN, which aims to accurately classify mental tasks in fNIRS-BCI. The CGAN-rIRN consists of two deep learning models, one of which is a convolution-based CGAN model to generate category-specific fNIRS single for data augmentation, and the other is an elaborate rIRN model for high-accuracy classification of multivariate time series based fNIRS signals. The CGAN-CON and rIRN models address the two challenges of insufficient fNIRS data for DLC training and accuracy improvement, respectively. The effectiveness and superiority of the CGAN-rIRN was validated by performing paradigm experiments. The results demonstrate that both CGAN-CON-based data augmentation and specifically-designed rIRN enhance the accuracy of MA and MS mental tasks. The CGAN-rIRN achieved a maximum average accuracy of 92% across all subjects, which was obtained using augmented data of equal size to the original data. Overall, these encouraging results demonstrate the potential of the CGAN-rIRN approach in improving the accuracy of mental tasks and provide a new perspective for real-time multi-class mental task classification in daily life and clinical applications. The approach can be used clinically for functional brain imaging, localization, disease diagnosis, neurorehabilitation, and child development studies.

Funding. National Natural Science Foundation of China (61575140, 62075156, 81871393, 81971656).

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. M. N. A. Khan and K. S. Hong, "Most favorable stimulation duration in the sensorimotor cortex for fNIRS-based BCI," *Biomed. Opt. Express* **12**(10), 5939–5954 (2021).
2. M. J. Khan and K.-S. Hong, "Passive BCI based on drowsiness detection: an fNIRS study," *Biomed. Opt. Express* **6**(10), 4063–4078 (2015).
3. E. Hernandez-Martin, F. Marcano, C. Modrono, N. Janssen, and J. L. Gonzalez-Mora, "Diffuse optical tomography to measure functional changes during motor tasks: a motor imagery study," *Biomed. Opt. Express* **11**(11), 6049–6067 (2020).
4. S. D. Power, A. Kushki, and T. Chau, "Automatic single-trial discrimination of mental arithmetic, mental singing and the no-control state from prefrontal activity: toward a three-state NIRS-BCI," *BMC Res. Notes* **5**(1), 141 (2012).
5. H. J. Hwang, J. H. Lim, D. W. Kim, and C. H. Im, "Evaluation of various mental task combinations for near-infrared spectroscopy-based brain-computer interfaces," *J. Biomed. Opt.* **19**(7), 077005 (2014).
6. S.-H. Yoo, S.-W. Woo, and Z. Amad, "Classification of three categories from prefrontal cortex using LSTM networks: fNIRS study," in *18th International Conference on Control, Automation and Systems (ICCAS), International Conference on Control Automation and Systems* (2018), 1141–1146.

7. T. Ma, S. Wang, Y. Xia, X. Zhu, J. Evans, Y. Sun, and S. He, "CNN-based classification of fNIRS signals in motor imagery BCI system," *J. Neural Eng.* **18**(5), 056019 (2021).
8. Y. Zhang, D.-Y. Liu, P.-R. Zhang, T.-N. Li, Z.-Y. Li, and F. Gao, "Combining robust level extraction and unsupervised adaptive classification for high-accuracy fNIRS-BCI: An evidence on single-trial differentiation between mentally arithmetic- and singing-tasks," *Front. Neurosci.* **16**, 938518 (2022).
9. S. D. Power, A. Kushki, and T. Chau, "Towards a system-paced near-infrared spectroscopy brain-computer interface: differentiating prefrontal activity due to mental arithmetic and mental singing from the no-control state," *J. Neural Eng.* **8**(6), 066004 (2011).
10. A. Janani, M. Sasikala, H. Chhabra, N. Shajil, and G. Venkatasubramanian, "Investigation of deep convolutional neural network for classification of motor imagery fNIRS signals for BCI applications," *Biomedical Signal Processing and Control* **62**, 102133 (2020).
11. K.-S. Hong, U. Ghafoor, and M. J. Khan, "Brain-machine interfaces using functional near-infrared spectroscopy: a review," *Artificial Life and Robotics* **25**(2), 204–218 (2020).
12. H. J. Hwang, H. Choi, J. Y. Kim, W. D. Chang, D. W. Kim, K. W. Kim, S. H. Jo, and C. H. Im, "Toward more intuitive brain-computer interfacing: classification of binary covert intentions using functional near-infrared spectroscopy," *J. Biomed. Opt.* **21**(9), 091303 (2016).
13. Y. Zhang, B. Wang, and F. Gao, "Real-time decoding for fNIRS-based brain computer interface using adaptive Gaussian mixture model classifier and Kalman estimator," in *Asia Communications and Photonics Conference (ACP), Asia Communications and Photonics Conference and Exhibition* (IEEE, 2018).
14. T. Trakoolwilaiwan, B. Behboodi, J. Lee, K. Kim, and J.-W. Choi, "Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain-computer interface: three-class classification of rest, right-, and left-hand motor execution," *Neurophotonics* **5**(01), 1 (2017).
15. S. Hiwa, K. Hanawa, R. Tamura, K. Hachisuka, and T. Hiroyasu, "Analyzing brain functions by subject classification of functional near-infrared spectroscopy data using convolutional neural networks analysis," *Comput Intel Neurosci* **2016**, 1–9 (2016).
16. U. Asgher, K. Khalil, M. J. Khan, R. Ahmad, S. I. Butt, Y. Ayaz, N. Naseer, and S. Nazir, "Enhanced accuracy for multiclass mental workload detection using long short-term memory for brain-computer interface," *Front. Neurosci.* **14**, 584 (2020).
17. C. Lin, G. Zhao, Z. Yang, A. Yin, X. Wang, L. Guo, H. Chen, Z. Ma, L. Zhao, H. Luo, T. Wang, B. Ding, X. Pang, and Q. Chen, "CIR-Net: automatic classification of human chromosome based on inception-resnet architecture," *IEEE/ACM Trans. Comput. Biol. and Bioinf.* **19**, 1 (2020).
18. M. Saadati, J. Nelson, and H. Ayaz, "Mental workload classification from spatial representation of fNIRS recordings using convolutional neural networks," *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, Pittsburgh, PA, USA (2019).
19. H. Ghonchi, M. Fateh, V. Abolghasemi, S. Ferdowsi, and M. Rezvani, "Deep recurrent-convolutional neural network for classification of simultaneous EEG-fNIRS signals," *IET signal process.* **14**(3), 142–153 (2020).
20. K. Khalil, U. Asgher, and Y. Ayaz, "Novel fNIRS study on homogeneous symmetric feature-based transfer learning for brain-computer interface," *Sci. Rep.* **12**(1), 3198 (2022).
21. X.-Z. Zhang, Z.-Z. Liu, J.-F. Jiang, K. Liu, X.-J. Fan, B.-Y. Yang, M. Peng, G. L. Chen, and T. G. Liu, "Data augmentation of optical time series signals for small samples," *Appl. Opt.* **59**(28), 8848–8855 (2020).
22. F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," *Int Conf 3d Vision*, 565–571 (2016).
23. J. Dinares-Ferran, R. Ortner, C. Guger, and J. Sole-Casals, "a new method to generate artificial frames using the empirical mode decomposition for an EEG-based motor imagery BCI," *Front. Neurosci.* **12**, 308 (2018).
24. T. Nagasawa, T. Sato, I. Nambu, and Y. Wada, "fNIRS-GANs: data augmentation using generative adversarial networks for classifying motor tasks from functional near-infrared spectroscopy," *J. Neural Eng.* **17**(1), 016068 (2020).
25. S. D. Wickramaratne and M. S. Mahmud, "Conditional-GAN based data augmentation for deep learning task classifier improvement using fNIRS Data," *Front. Big Data* **4**, 659146 (2021).
26. L. Duan, Z.-P. Zhao, Y.-L. Lin, X.-Y. Wu, Y.-J. Luo, and P.-F. Xu, "Wavelet-based method for removing global physiological noise in functional near-infrared spectroscopy," *Biomed. Opt. Express* **9**(8), 3805–3820 (2018).
27. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *28th Conference on Neural Information Processing Systems (NIPS), Advances in Neural Information Processing Systems* (2014), 2672–2680.
28. J. Gauthier, "Conditional generative adversarial nets for convolutional face generation.,", in *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition*, MIT Press: Winter Semester Cambridge, MA 2014, 2.
29. D. P. Kingma and J. L. Ba, "Adam: a method for stochastic optimization," *Computer Science* (2014).
30. J. Jeppkoech, D. M. Mugo, B. K. Kenduiywo, and E. C. Too, "The effect of adaptive learning rate on the accuracy of neural networks," *International Journal of Advanced Computer Science and Applications* **12**(8), 736–751 (2021).
31. W.-T. Chen, X. Wang, B.-Y. Wang, Y.-H. Wang, Y.-Q. Zhang, H.-J. Zhao, and F. Gao, "Lock-in-photon-counting-based highly-sensitive and large-dynamic imaging system for continuous-wave diffuse optical tomography," *Biomed. Opt. Express* **7**(2), 499–511 (2016).

32. H.-J. Zhao, F. Gao, Y. Tanikawa, Y. Onodera, M. Ohmi, M. Haruna, and Y. Yamada, "Imaging of in vitro chicken leg using time-resolved near-infrared optical tomography," *Phys. Med. Biol.* **47**(11), 3101979 (2002).
33. Y. Gao, H. Chao, L. Cavuoto, P. Yan, U. Kruger, J. E. Norfleet, B. A. Makled, S. Schwartzberg, S. De, and X. Intes, "Deep learning-based motion artifact removal in functional near-infrared spectroscopy," *Neurophoton.* **9**(04), 041406 (2022).
34. H. Ding, Q. Lu, H. Gao, and Z. Peng, "Non-invasive prediction of hemoglobin levels by principal component and back propagation artificial neural network," *Biomed. Opt. Express* **5**(4), 1145–1152 (2014).
35. W. Li, J. Chen, Z. Wang, Z. Shen, C. Ma, and X. Cui, "IFL-GAN: Improved Federated Learning Generative Adversarial Network With Maximum Mean Discrepancy Model Aggregation," *IEEE Transactions on Neural Networks and Learning Systems* (2022).
36. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. on Image Process.* **13**(4), 600–612 (2004).
37. B. Yang, Y. Lei, F. Jia, N. Li, and Z. Du, "A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines," *IEEE Trans. Ind. Electron.* **67**(11), 9747–9757 (2020).
38. K. M. He, X. Y. Zhang, S. Q. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc CVPR IEEE* 770–778 (2016).
39. A. Hamann and N. Carstengerdes, "Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement," *Sci Rep* **13**(1), 4738 (2023).
40. R. D. S. Raizada and A. C. Connolly, "What makes different people's representations alike: neural similarity space solves the problem of across-subject fMRI decoding," *Journal of Cognitive Neuroscience* **24**(4), 868–877 (2012).
41. L. B. Jin and E. Y. Kim, "Interpretable cross-subject EEG-based emotion recognition using channel-wise features," *Sensors* **20**(23), 6719 (2020).
42. O. Y. Kwon, M. H. Lee, C. T. Guan, and S. W. Lee, "Subject-independent brain-computer interfaces based on deep convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems* **31**(10), 3839–3852 (2020).
43. H. G. Wen, J. X. Shi, W. Chen, and Z. M. Liu, "Transferring and generalizing deep-learning-based neural encoding models across subjects," *Neuroimage* **176**, 152–163 (2018).
44. S. Weyand, L. Schudlo, K. Takehara-Nishiuchi, and T. Chau, "Usability and performance-informed selection of personalized mental tasks for an online near-infrared spectroscopy brain-computer interface," *Neurophotonics* **2**(2), 025001 (2015).